EQC for Seasonal Forecasts C3S_51 Lot 3

Climate Change

Francisco Doblas-Reyes[1,2]

1 Barcelona Supercomputing Center (BSC)
2 Institució Catalana de Recerca i Estudis Avançats (ICREA)

Copernicus Europe's eyes on Earth

European Commission

Climate Change

Challenges of the EQC for seasonal

- **Address adaptation**: it must provide information for all kind of services including adaptation, taking advantage that many users are already familiar with the climate-change problem.
- **Provide consistency**: it must build trust, ensuring a high degree of coherence across products, underlying data sets, processing methods, communication, context, etc.
- **Provide innovation**: it should transfer recent developments from research to operations to answer real-world issues.
- **Address efficiency**: the EQC information should be timely, e.g. respond to users' queries with a delay as short as possible, which imposes conditions on the algorithms considered.
- **Define the target**: data, products, communication, etc.

Climate Change

Lot 3: QA4Seas

QA4Seas aims at **developing a strategy for the evaluation and quality control (EQC) of the multi-model seasonal forecasts provided by the Copernicus Climate Change Service (C3S) to respond to the needs identified among a wide range of stakeholders**.

To achieve the objective the consortium will:

- Consider the evaluation of multi-faceted quality aspects
- Be user driven with a two-stage consultation process (coordinated with other lots)
- Formulate requirements to the CDS to address user requirements
- Perform a gap analysis of the current information available to users
- Develop a framework and a prototype of the EQC system

Copernicus Europe's eyes on Earth

European Commission

Climate Change

QA for seasonal forecasts

Reference points for quality assurance (QA):

- A prediction has no real value without an estimate of its quality based on past performance.
- QA is multifaceted, no single metric fully characterises the quality of a system or allows to single-out the best forecast system.
- It addresses administrative (tracks the system evolution), scientific (skill and reliability), or socio-economic (users' requirements) questions.
- QA of the European multi-model is not readily available, contrary to the Asian and North American one.
- **Climate service providers should consider the consequences of their actions for those who may be affected by the products**.
- **Climate service products should be open to scrutiny and comparison**.

Copernicus Europe's eyes on Earth

European Commission

Climate Change

QA for seasonal forecasts

Partner	Nature	Role	Effort (PM)
BSC-CNS (ES)	Main contractor	Coordination, data inventory and EQC framework and prototype	88
Univ. Leeds (UK)	Subcontractor	Assess user requirements	13.92
Meteoswiss (CH)	Subcontractor	Scientific quality assessment and gap analysis	23
Predictia (ES)	Subcontractor	CDS requirements and development of the prototype	28
Univ. Exeter (UK)	Subcontractor	Expert statistical advice	3.72
IFCA-CSIC (ES)	Subcontractor	Downscaling	14

Copernicus Europe's eyes on Earth

European Commission

Climate Change

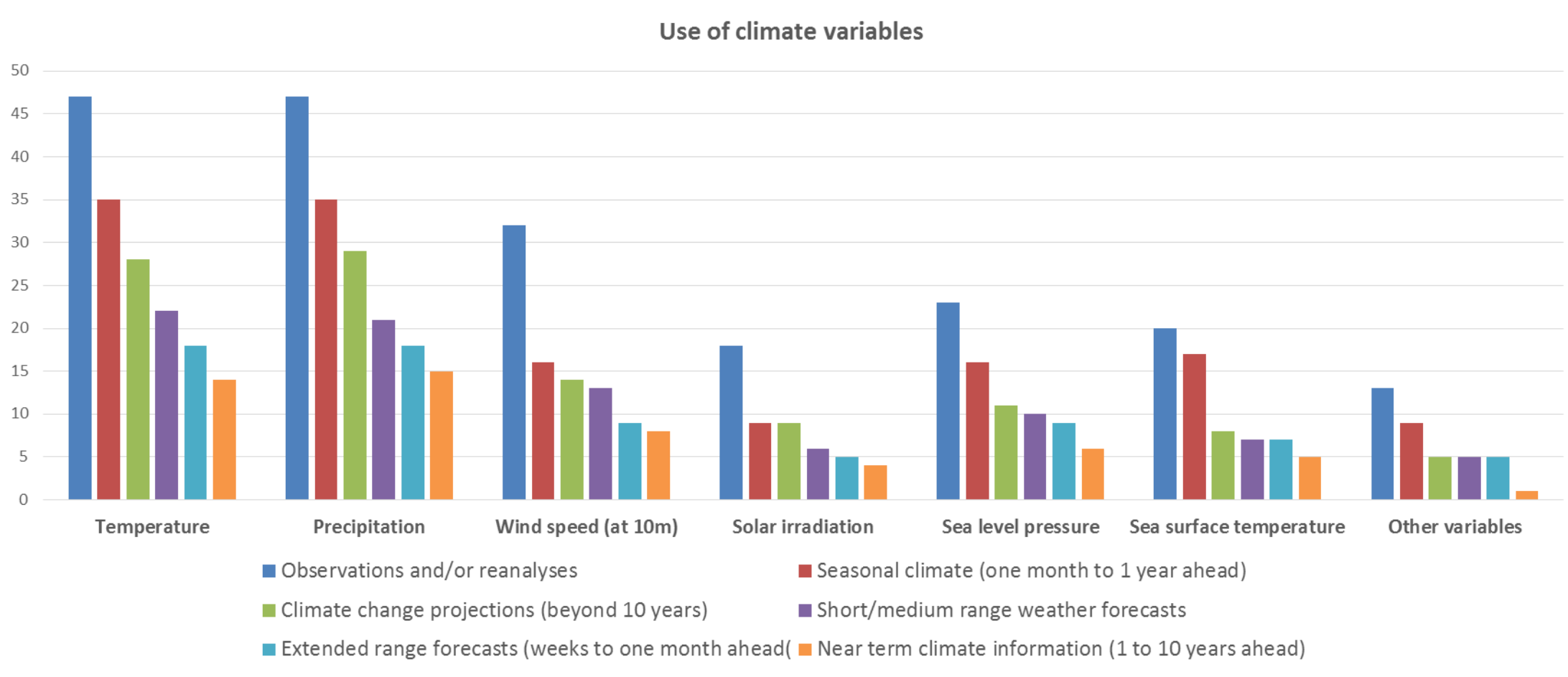
Users' requests

Results from a survey with 73 respondents that use seasonal forecast information, with a large majority of NMHSs.

- **57% responses from NMHS**, 15% research institutes and 11% private companies.
- Majority of responses from **climate scientists** and **researchers**.
- The large majority use meteorological/climatological data in their work.
- Sources of seasonal forecasts (SF) are **ECMWF** (n=31) and **NMHS** (n=26).
- Tend to use SF from **multiple providers** (n=30) or single provider (n=13).
- Accessibility through manual download (n=25), visual inspection (n=22) and/or automated download of SF (n=18).
- Formats used: NetCDF (n=21), GRIB (n=17), ASCII (n=13).
- Additional information available from SECTEUR.

Climate Change

Users' requests

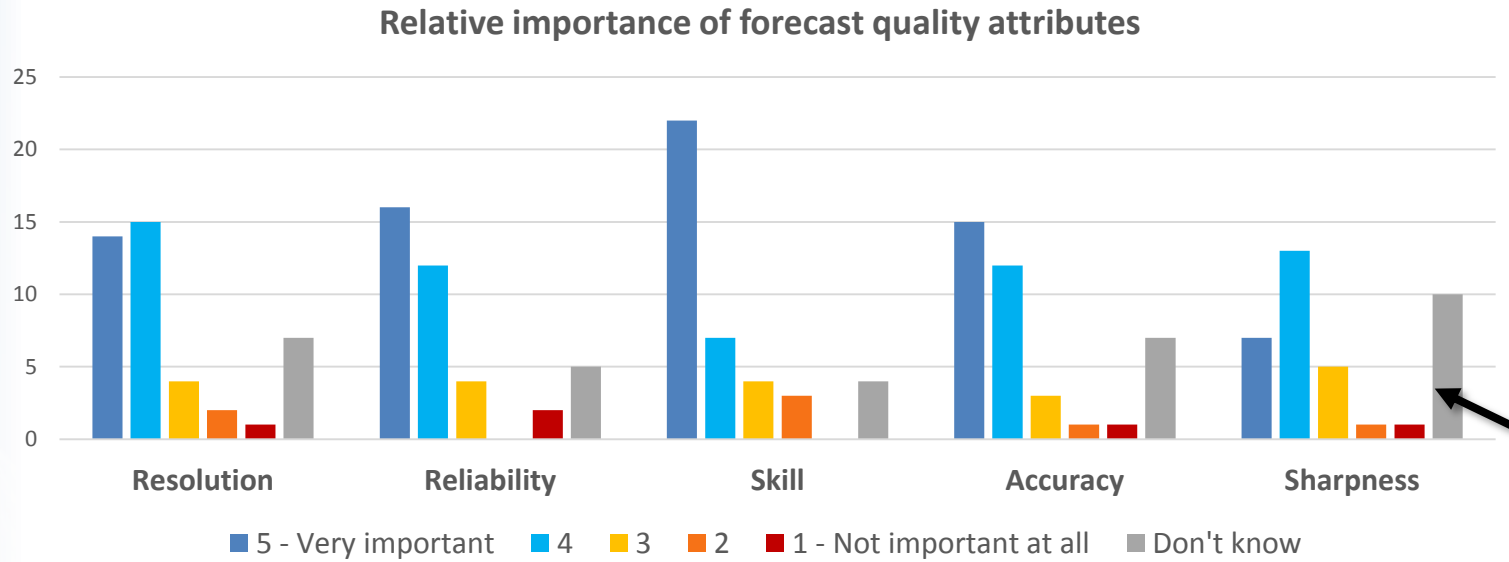


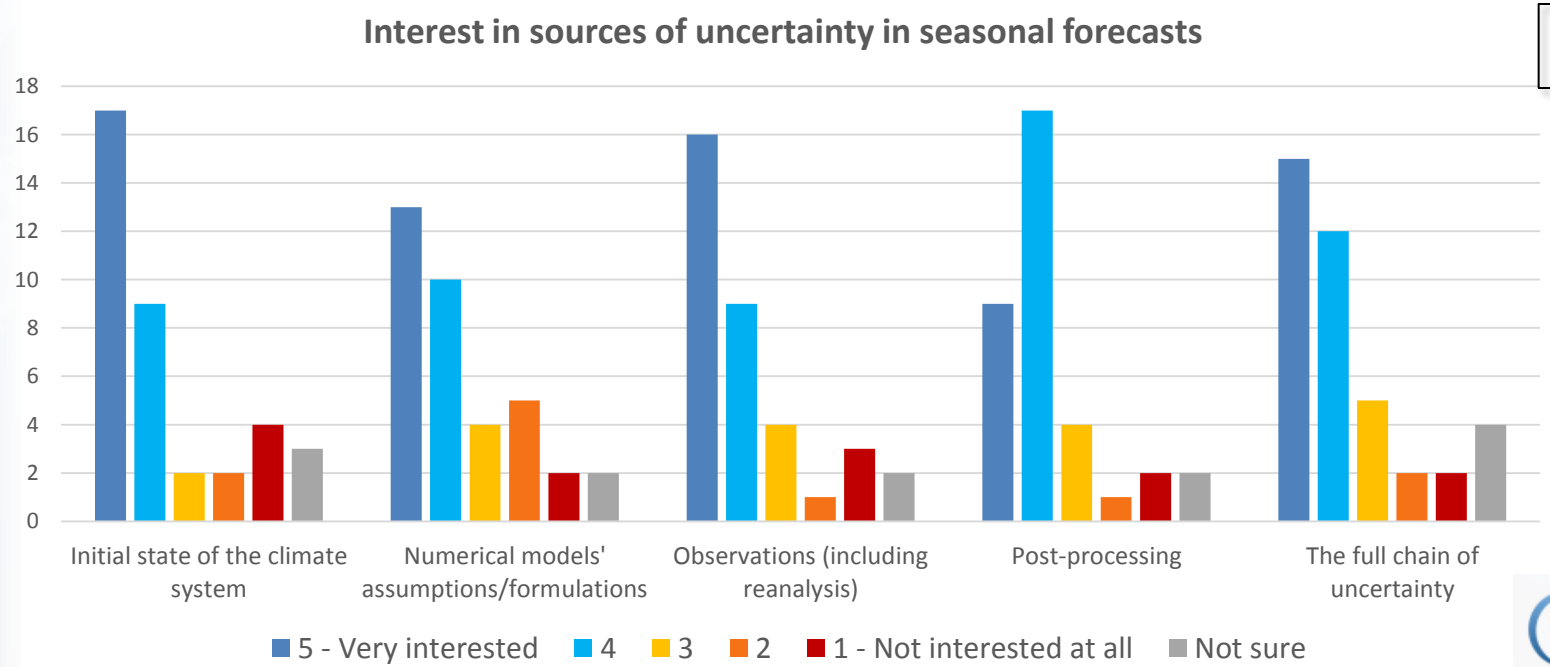
Users tend not to employ derived quantities or indices (n=27); those who do tend to use HDD and heavy precipitation/rainfall index (n=17 each), consecutive dry days (n=14) and cooling degree days (n=13).

M. Soares, A. Taylor (Univ. Leeds)

Users' requests

Climate Change





Note the grey bar

M. Soares, A. Taylor (Univ. Leeds)

Climate Change

Users' requests

A series of **interviews** is now being conducted:

- Aim to provide a better understanding of the requirements for evaluation and quality control regarding seasonal forecasts within and across large organisations.
- Around 12 interviews with key individuals, advisory board plus a few selected ones (e.g. IRI, ECMWF).
- Interview protocol developed from the survey experience to ensure consistency across the methods applied.
- Opportunity to introduce the preliminary list of functional requirements for the EQC prototype with the interviews: follow up and feedback after the interviews.
- Analysis of both survey and interviews expected at the end of the summer.

Climate Change

Data and metadata

Once the data are identified, a critical question is which of the available data will be used as the data use will depend on, among other things, the quality information.

- Identify the **data available**, both observations and seasonal forecasts (tables available for ocean and atmosphere).
- Assess (and define whenever necessary) all necessary **metadata** (defined for seasonal forecasts).
- Inadequate sources for the user's application to be pointed at (e.g. discard "bad" models or inadequate references).
- The sources retained should be diverse enough to still adequately represent all uncertainties (e.g. has the solution enough spread?)

What does "bad" mean?

Copernicus Europe's eyes on Earth

European Commission

Climate Change

Scientific assessment

A thorough assessment of the existing C3S seasonal forecasts from C3S_433 has been done:

- More than 30,000 plots made available via a Shiny app.
- An assessment of the results is being carried out, feeding the interviews at the same time. It will be made available as a paper.

The building blocks

Climate Change

Existing R packages are used for both the preliminary assessment and the prototype. New elements (inference assessment, metadata propagation, etc.) are built in to address the C3S needs.

SCORES

SpecsVerification
- Probabilistic and deterministic scores
- Works on [time x members] arrays

easyVerification
- Applies SpecsVerification scores to arrays of any dimensions, multi-core
- Probabilistic and deterministic scores

FRAMEWORKS

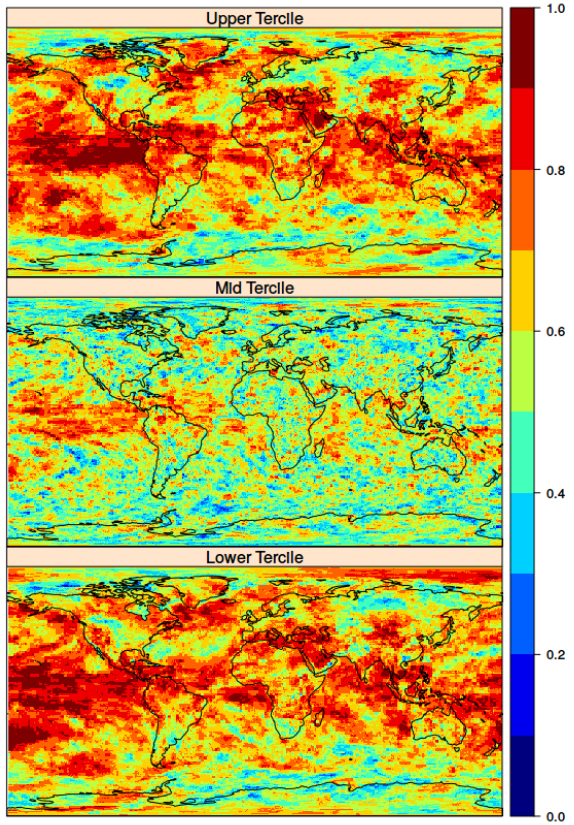
downscaleR + loadeR
- Data retrieval and homogenization
- Bias adjustment, modes, downscaling
- Probabilistic and deterministic scores
- Visualisation of data and results

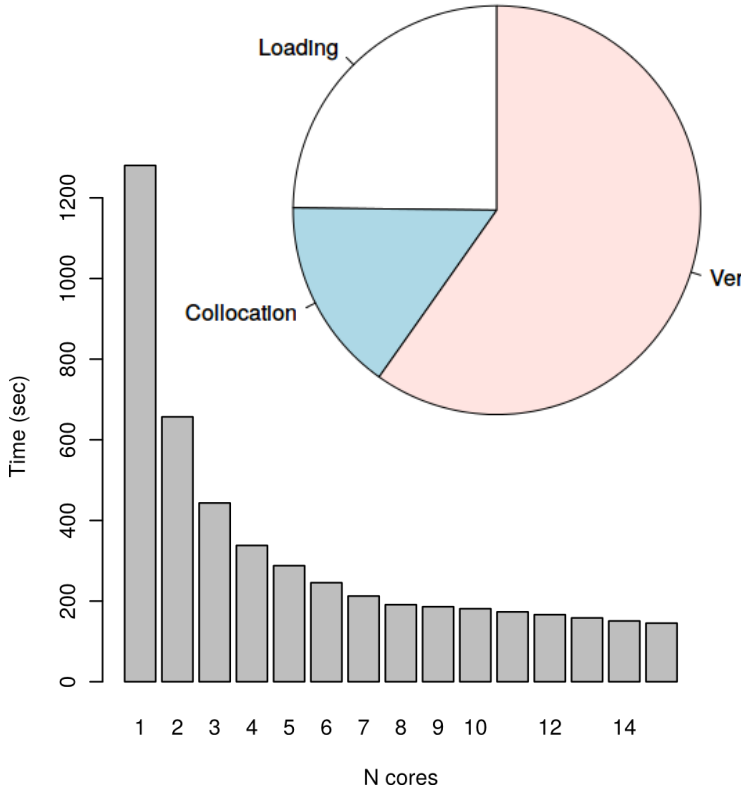
s2dverification
- Data retrieval and homogenization
- Bias adjustment, filtering, modes
- Probabilistic and deterministic scores
- Visualisation of data and results

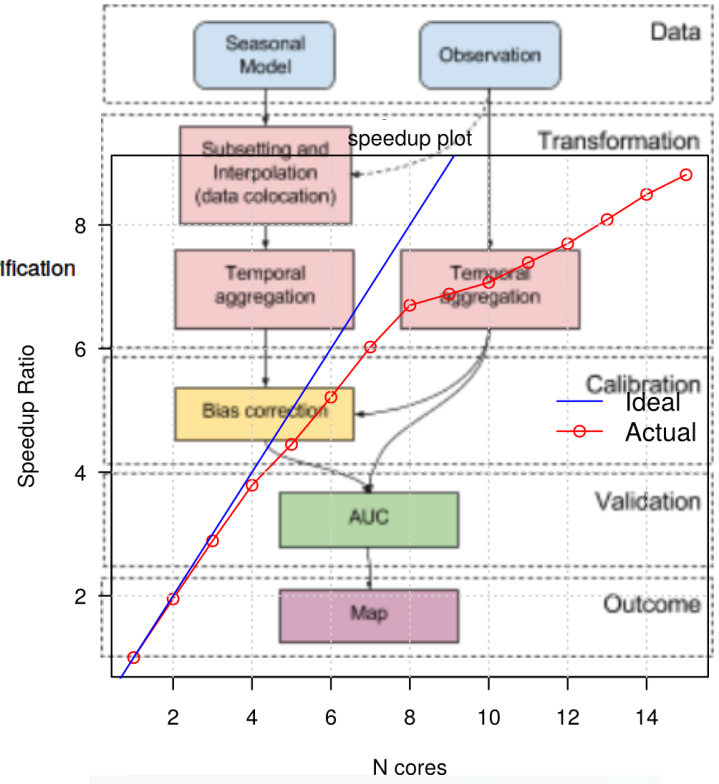
Climate Change

Gap analysis: links to the CDS

Computing performance is key: example of multi-model seasonal verification, with performance analysis of a ROC area estimate using loadeR, SpecsVerification and easyVerification.



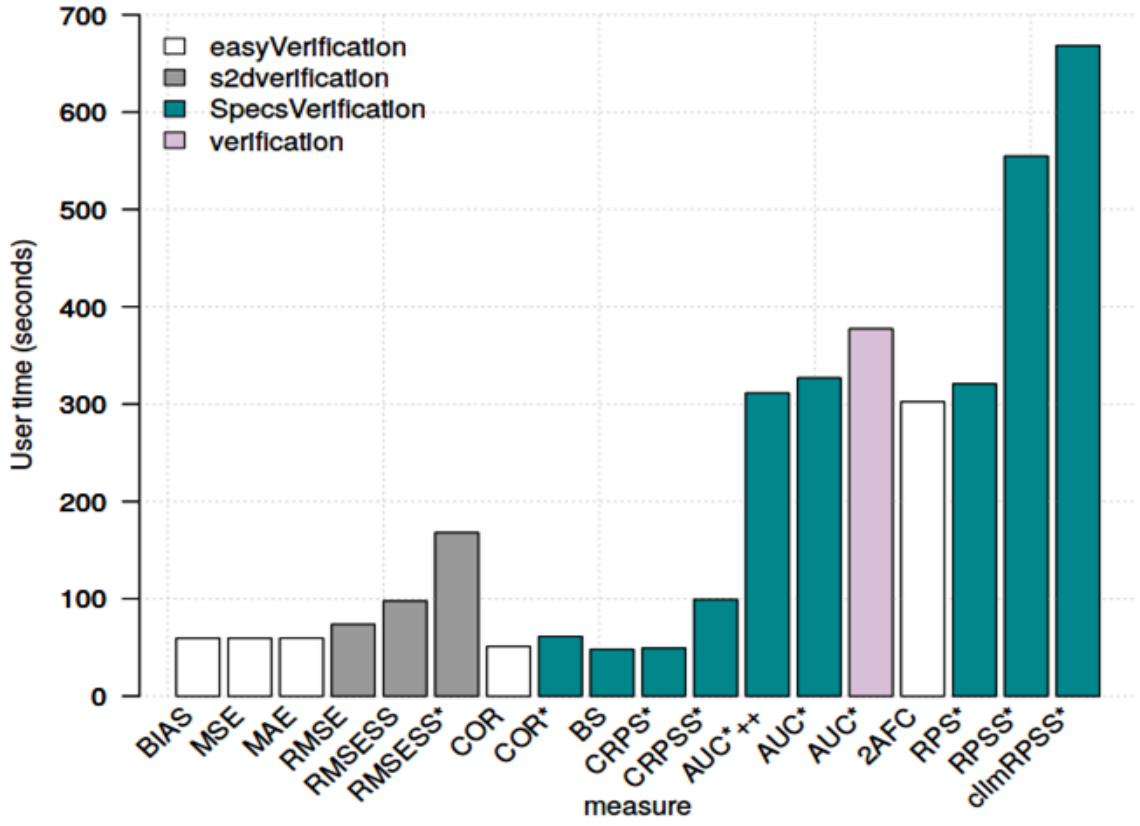




J. Bedia, D. San Martín (Predictia)

Gap analysis: links to the CDS

Computing performance is key: performance and efficiency (parallel scalability), which are evaluated in controlled environments, are part of the contract KPIs.



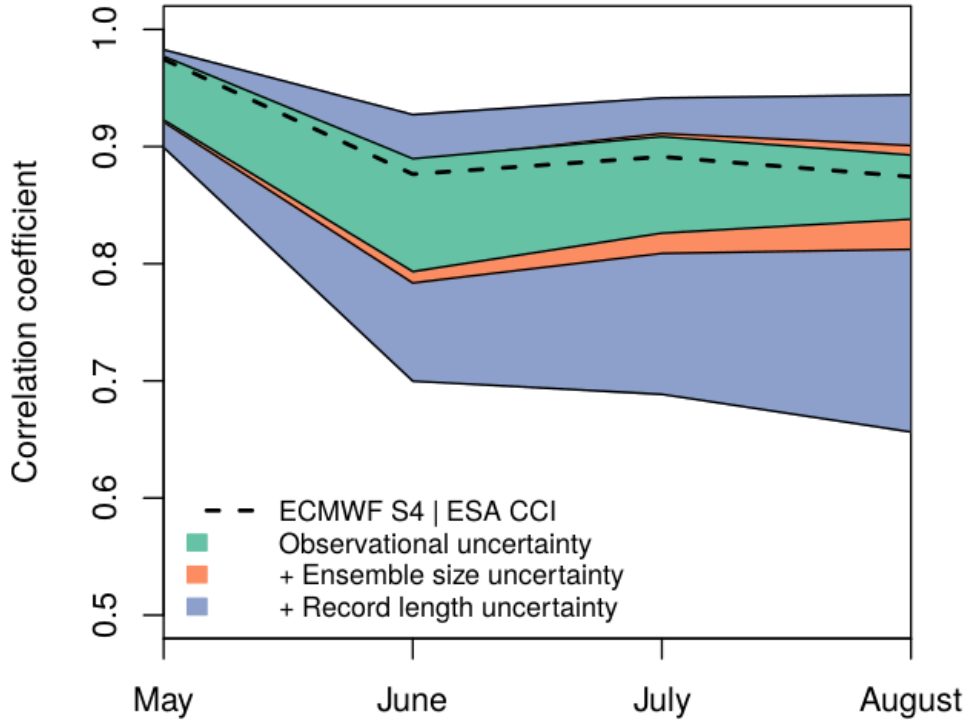
J. Bedia, D. San Martín (Predictia)

Climate Change

Gap analysis: observational uncertainty

Making the most of C3S: bringing in observational and reference uncertainty in the EQC process.

Niño 3.4 SST correlation of the ensemble mean for EC-Earth3.1 (T511/ORCA025) predictions with ERAInt and GLORYS2v1 ics, and BSC sea-ice reconstruction started every May over 1993-2009



O. Bellprat (BSC)

CDS requirements and EQC framework

How to identify data/products to ensure a minimum quality?

- **Reproducibility**: ability of an entire process to be duplicated.
- **Traceability**: ability to verify the history, location or application of an item by means of documented recorded identification.
- **Context**: built-in documentation is required.

Metadata are essential to support these features:

- What is the input data and their temporal/spatial resolution?
- What are the transformation steps involved? In what order?

More information than traditional metadata is required to create context, involving a **domain-specific conceptual description**:

- What measures AUC?
- How was it computed?

CDS requirements and EQC framework

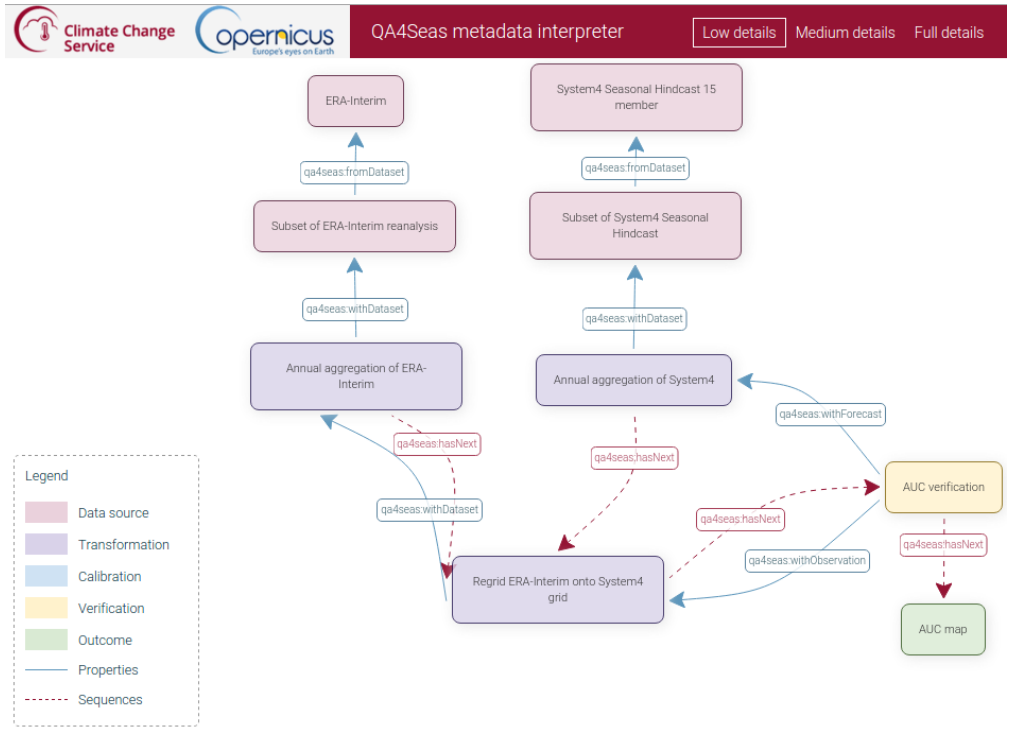
How to identify data/products to ensure a minimum quality?

- Generalised **metadata and provenance** information are key elements of all the components of the service.
- Two **approaches** for product provenance (but there are more) are under discussion: S-PROV and Resource Description Framework (RDF). They are not mutually exclusive.

CDS requirements and EQC framework

How to identify data/products to ensure a minimum quality?

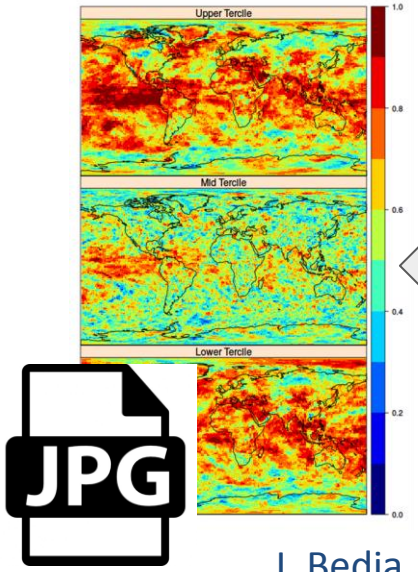
- The RDF-based approach aims at the reproducibility of objects (NetCDF file, image) with human and machine-readable solution.
- It uses a semantic metadata model that builds the vocabularies on existing initiatives (e.g. VALUE for downscaling).



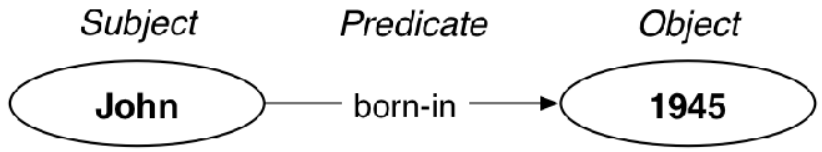
J. Bedia, D. San Martín (PREDICTIA)

How to identify data/products to ensure a minimum quality?

- The RDF-based approach uses World Wide Web Consortium (W3C) standard (set on 10 February 2004) with a general method for conceptual description of information (origin in the semantic web) and with graph representations.
- Based on triples: subject-predicate-object.



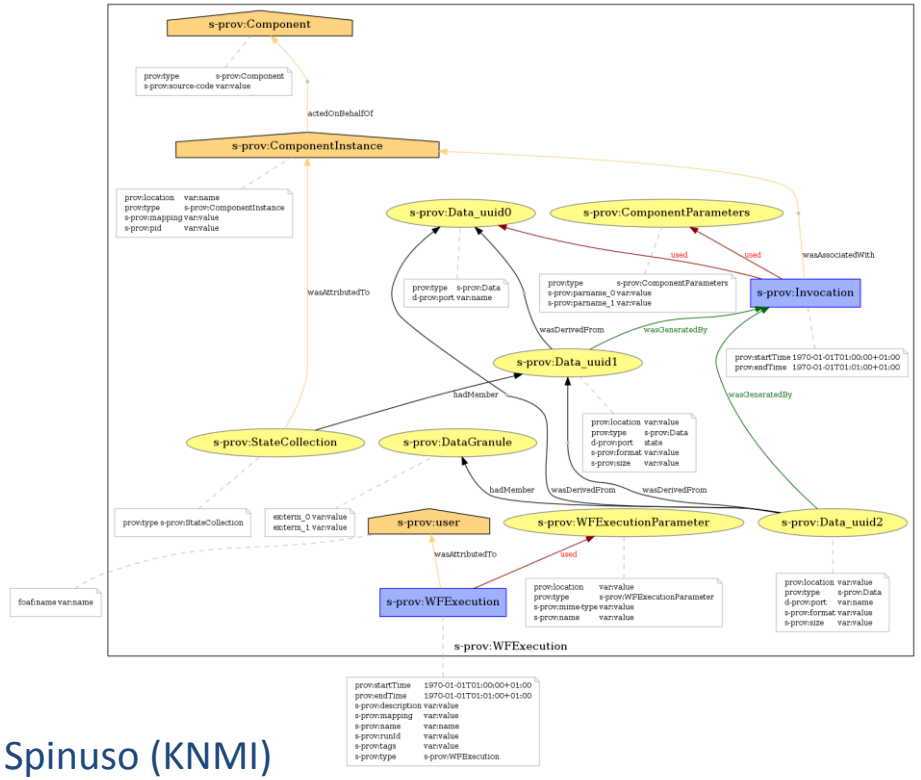
The metadata is delivered within the image



CDS requirements and EQC framework

How to identify data/products to ensure a minimum quality?

- S-PROV represents the relationships occurring between the players of a data-intensive computation in a scientific domain.
- Uses W3C PROV as data model and S-PROVFlow for reproducibility as a service.



CDS requirements and EQC framework

List of issues to discuss with the CDS:

- Define the common data model and the best way to access the datasets.
- Define the curation of elements other than raw data.
- Feedback on the roadmap for the design of the prototype, in particular the most efficient way to load the data.
- Relevance of the computational and memory efficiency work.
- Provenance challenges: define the level of granularity to describe the objects, inform about and display different levels of abstraction, governance.

Climate Change

Summary

- EQC is user driven, but user requirements are a moving target that also depends on the amount of information they receive.
- EQC information is not neutral, precise definitions are necessary, documentation (context, provenance) is key.
- Data inventories help identify gaps: relevance, metadata, etc.
- Existing packages are an invaluable source of solutions, and are considered within a framework that addresses adequacy, efficiency and provenance.
- Handling metadata and provenance information requires a generic, common approach (e.g. common data model) for all the EQC (and other components) work in C3S.